Title:

Reliability of two techniques for assessing cerebral iron deposits from structural MRI

Authors:

Maria C. Valdés Hernández[1,2], PhD; Tina H. Jeong[3]; Catherine Murray[2,4], MA (Hons); Mark E. Bastin[1,2,5], DPhil; Francesca M. Chappell[1], PhD; Ian J. Deary[2,4], MB, ChB, PhD (Edin), FRCPsych, FRCPE, FBA, FRSE, FMedSci; and Joanna M. Wardlaw[1,2], MB ChB, MD, FRCP, FRCR, FMedSci

Authors Affiliations:

[1]SINAPSE Collaboration, SFC Brain Imaging Research Centre, Department of Clinical Neurosciences, University of Edinburgh, Edinburgh, UK

[2]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK

[3]School of Medicine and Veterinary Medicine, University of Edinburgh, Edinburgh, UK

[4]Department of Psychology, University of Edinburgh, Edinburgh, UK

[5]Medical and Radiological Sciences (Medical Physics), University of Edinburgh, Edinburgh, UK

Correspondence:

Dr. Maria C. Valdés Hernández; SFC Brain Imaging Research Centre. Image Analysis Lab. Division of Clinical Neurosciences. University of Edinburgh. Western General Hospital. Crewe Road, Edinburgh EH4 2XU. UK. Telephone: +44-131-537 3669, Fax: +44-131-332 5150, e-mail: mvhernan@staffmail.ed.ac.uk

Grant Support:

MVH is supported by The Disconnected Mind (www.disconnectedmind.ed.ac.uk) funded by Help the Aged and the UK Medical Research Council; THJ was funded by Carnegie Trust for the universities of Scotland; JMW is supported by the Scottish Funding Council (SFC) through the SINAPSE Collaboration (Scottish Imaging Network, A Platform for Scientific Excellence, www.sinapse.ac.uk). The SFC Brain Imaging Research Centre is supported by the SINAPSE Collaboration. MVH, MEB, IJD and JMW are members of The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative. Funding from the BBSRC, EPSRC, ESRC and MRC is gratefully acknowledged.

Running Title:

Two methods to assess iron deposits

Purpose: To test the reliability of two computational methods for segmenting cerebral iron deposits (IDs) in the ageing brain, given that its measurement in MRI is challenging due to the similar effect produced by other minerals, especially calcium, on T2*-weighted sequences.

Materials and Methods: T1-, T2*-weighted and FLAIR MR brain images obtained at 1.5T from 70 subjects in their early 70s who displayed a wide range of brain IDs were analyzed. The first segmentation method used a multispectral approach based on the fusion of two or more structural sequences registered and mapped in the red/green color space followed by Minimum Variance Quantization. The second method employed a combined thresholding, size and shape analysis using T2*-weighted images augmented with visual information from T1-weighted data.

Results: Both segmentation techniques had high intra- and inter-observer agreement (95 % CI = ± 57 voxels in a range from 0 to 1800), which decreased in subjects with significant microbleeds and/or IDs. However, the thresholding method was more observer dependent in identifying microbleeds and IDs boundaries than the multispectral approach.

Conclusion: Both techniques proved to be in agreement and have good intra- and inter-observer reliability. However, they have limitations, specifically with regard to automation and observer independence, so further work is required to develop fully user-independent methods of identifying cerebral IDs.

Key Words: iron deposits; MRI; thresholding; multispectral segmentation; reliability

Iron is typically stored in the body as a soluble oxyhydroxide in the protein ferritin, which controls the release of iron and prevents toxic damage to tissues. However, an insoluble form of iron oxyhydroxide, haemosiderin, can accumulate in tissues in some diseases, leading to organ damage. These iron deposits (IDs) occur as haemosiderin in the brain after intraparenchymal haemorrhage, on the surface of the brain following subarachnoid or subdural haemorrhage (superficial siderosis), and in microhaemorrhages in brain tissue. In addition, iron may be deposited in the walls of small blood vessels, for example in the perforating arterioles as they enter the brain substance in the inferior part of the putamen (1).

Cerebral IDs are of increasing interest due to their association with some diseases (2-4), mood disorders (5,6) and cognition (7). In addition to iron, other metals and compounds including calcium and manganese may be deposited in the brain in pathological states. For example, in the case of putaminal IDs, pathology studies show that small arteriolar mineralization has staining properties of both iron and calcium (1,4). These paramagnetic substances can affect the longitudinal (T1) and transverse (T2/T2*) relaxation times of mobile water protons predominantly through the outer sphere mechanism (8), with iron producing a reduction in T2/T2* relaxation time with little effect on T1, and calcium and manganese also reducing T2/T2* but in addition altering T1 depending on the pathological state (8). A shortening of T2/T2* relaxation time produces hypointensity on T2-weighted (T2W) and T2*-weighted (T2*W) MRI, while altering T1 produces either hyperintensity or hypointensity on T1-weighted (T1W) MRI. Since calcium is hypointense on T2*W and T1W MRI and iron is hypointense on T2*W but not T1W, this enables differentiation of these two minerals using structural brain MRI.

The assessment of IDs by MRI offers novel and useful applications for diagnosis, longitudinal monitoring, and testing of new therapies for brain disorders. Several studies assessed brain microbleeds (BMBs) (9) and hemosiderin deposits visually (2,3,6,10), and attempts have

been made to segment these regions using computational image processing methods. For example, a Fourier-based method to differentiate iron and calcium was proposed by T. Freeman ("MRI analysis: a study of uncertainty"; http://medicalphysicsweb.org/cws/article/research/38012) who used a Green's function to calculate magnetic susceptibilities from given field distributions. More recently, the use of phase imaging has been investigated in an attempt to assess areas with IDs (11). However, both these methods are emerging techniques whose accuracy and utility are still to be determined.

In this paper, we implemented and tested two techniques for segmenting brain IDs from a cohort of relatively healthy subjects in their early 70s using structural MRI data. These methods have previously been used to segment white matter hyperintensities (WMHs) in several studies of normal ageing and have been found to be reliable (12-14). The first method, which is considered to be the gold standard technique, is conventional thresholding followed by manual editing. The second is a multispectral technique which has been shown to segment brain tissue and WMHs accurately in ageing and pathological brain MRI data (15). Here we examine their intra- and inter-observer repeatability and reliability in segmenting IDs for research and clinical use.

MATERIALS AND METHODS

MRI Data

Structural T1W, T2*W and FLAIR MRI data were obtained from 325 community dwelling older subjects, born in 1936 and scanned at the age of 70 to 71 years, who were participating in a study of cognitive ageing. MRI data were acquired on a GE 1.5T HDX clinical scanner, and all subjects gave formal written consent. Table 1 displays the sequence parameters used to scan the participants; the slice location was contiguous in all cases. All sequences were registered to the

T2W scan using FLIRT (FMRIB, Oxford, UK; http://www.fmrib.ox.ac.uk) giving a final image matrix of 256 × 256.

Visual Rating of Iron Deposits

IDs were visually assessed in 325 participants by an experienced image analyst with more than 15 years experience of using the following three visual rating scales. The Brain Observer MicroBleed Scale (BOMBS) minimizes observer variation in the assessment of BMBs in clinical practice (16). Since there is no established, validated visual scale for categorizing the putaminal IDs, we developed a simple visual rating scale to categorize the putaminal deposits, the 'Putaminal ID Visual Rating Scale'. This rates putaminal IDs from none to medium-high based on comparison with four standard cases (Figure 1).

We also developed a General Visual Rating Scale to quantify all IDs, i.e. microbleeds, superficial siderosis, old parenchymal haemorrhages and putaminal IDs. This scale encompassed all deposits considered to represent iron, regardless of location, i.e. on any slice on which the brain was visible, and included all sizes from small brain microbleeds (BMBs) to large iron deposits in the basal ganglia and remnants of old haemorrhages, regardless of shape, e.g. round BMBs to long, narrow areas of cortex siderosis. IDs were graded as: 0 (none), i.e. absence of any visible iron deposition; 1 (mild), i.e. equal or less than 5 BMBs or deposits whose extent was estimated to be less than 50 mm^3; 2 (moderate), i.e. from 6 to 30 BMBs or deposits whose extent was estimated to be between 50 mm^3 and 200 mm^3; and 3 (severe), i.e. more than 30 BMBs and/or deposits whose extent was estimated to be more than 200 mm^3. As a guide to estimate the volume of IDs, one BMB had, on average, a volume of 8 mm^3. These visual rating scales provided a simple method for describing BMB and IDs in our population.

Scan Selection

To evaluate the two segmentation methods, we randomly selected a subset of 70 subjects from the original 325 participants who displayed the full range of IDs based on the above visual rating scales: from none to significant BMBs and other IDs, e.g. deposits in the basal ganglia, midbrain, old haemorrhages, superficial siderosis, etc. We were careful to ensure that there were equal proportions of subjects in each iron load category and that all types of IDs were included (Table 2). The negative values of the kurtosis of the sample on both General and Putaminal Visual Rating Scales indicate that all grades of IDs on each scale were represented approximately equally throughout the sample.

Imaging Features

Figure 2a shows an axial slice from T2*W and FLAIR scans in a representative subject. This shows that the main signal intensity level groups in each sequence are associated with cerebrospinal fluid (CSF), which produces the highest intensities in T2*W and the lowest in FLAIR, and WMHs which result in medium to high signal intensities in T2*W and high intensities in FLAIR. Normal-appearing brain tissues, i.e. grey matter and white matter, are in the medium range of intensities in both sequences. As both segmentation techniques rely on differences in signal intensity levels to differentiate features of interest and therefore might be biased by brains with high WMH load or atrophy that would increase the amount of CSF signal, we first tested whether there was any association between background signal intensities from CSF or WMHs and the volume of iron-containing tissue in all 325 subjects. CSF and WMHs were quantified using the MCMxxxVI technique following an extensive validation process (15).

Differentiation of Iron and Other Minerals

T1W scans were used to differentiate areas with calcium deposits, i.e. signal change compared with no signal change for IDs. As described below, T2*W sequences were used in isolation in the thresholding technique and combined with FLAIR in MCMxxxVI to segment regions with high iron content.

Two trained observers, blind to each other's findings and to the rating scores, applied the two segmentation methods to measure ID volumes in the 70 subjects. One observer repeated the analysis with MCMxxxVI, blind to other results and without referring to the original T2*W sequence, and tested the effect that simultaneous visual assessment of T2*W images had on the segmentation obtained using this method.

Iron Volume Measurement: Multispectral Segmentation Method (MCMxxxVI)

In MCMxxxVI, T2*W and FLAIR sequences were registered using affine linear registration in FLIRT (17). The intensity values of the scans were adjusted to optimize their contrast prior to fusing to obtain a volume in the RG color plane (Figure 2 b). This step guarantees that when the registered images are transformed into the hue, saturation and value (HSV) color space (18) with an angle of 120˚, i.e. red and green colors, the features to be segmented are far enough from the V axis, S = 0 for any value of H, that the model which describes this transformation will not become undefined. A brain mask was then obtained from the T2*W data using the Object Extraction Tool in Analyze 8.1 (www.analyzedirect.com) which applies thresholding, morphological erosion, dilation, and region growing steps to separate the brain from the skull. To segment and quantify the volume of IDs, Minimum Variance Quantization (MVQ) was applied using the implementation performed in the MATLAB (www.mathworks.com) function 'rgb2ind' which converted the fused RG T2*W and FLAIR scans into clustered sequences in the same RG

color space. In (15) we found that 32 clusters was the optimum choice for achieving a good segmentation, and mapped the 32 clusters in a normalized graph of the RG space. We determined the clusters in the range of green that best discriminated the hemosiderin areas through interactive sampling. The segmentation was done automatically, followed by a manual removal of false positives where required. (Note MCMxxxVI is not designed for counting features, i.e. the number of IDs, and was therefore used only for measuring ID volume. An additional processing step for object recognition using morphological operations would be needed for MCMxxxVI to be used to count objects.)

Iron Volume Measurement: Thresholding Method

After extracting the brain using the T2*W volume as described above, bias field correction was employed to minimize the effect of signal intensity drop-off near the edges of the T2*W images using the Guillemaud filter (19). Next a slice was selected with significant BMBs or putaminal IDs, ideally with a variety of shapes and intensities, to allow the intensity threshold to be adjusted, between zero and less than half of the maximum intensity value, for optimal segmentation of areas with low intensity. An estimated maximum and minimum size of the hypointense “objects” was then adjusted interactively. The hypointense areas on T2*W images that satisfied the requirements of maximum size, circularity and specified threshold range consistent with IDs were extracted using the ‘Object Counter’ module in Analyze 8.1. Once the T2*W segmentation was complete, the slice in the registered T1W volume corresponding with that used in the T2*W segmentation was visually examined to identify areas where calcium dominated, i.e. signal change on T1W and hypointense on T2*W (8,15). These calcium-dominant areas were removed manually as well as any false positives, e.g. blood vessels and choroid plexus.

Statistical Analysis

SPSS 14.0 (SPSS Inc, Chicago, Ill, USA) was used to perform the statistical analyses. All variables were assessed for normality using the Kolmogorov-Smirnov test. The inter- and intra-observer repeatability of both segmentation methods was assessed using Bland-Altman analysis (20) in the sample of 70 subjects.

RESULTS

The effect of CSF and WMH volume on iron load was assessed using linear regression on the full cohort. Neither WMHs nor CSF volume were significantly associated with iron load as determined using the visual rating scales ($p > 0.18$) (Table 3, Figure 3). This indicates that the following results are not influenced by either atrophy or WMHs.

Counting the Number of IDs

The mean inter-observer difference obtained by the T2*W thresholding method was less than 4 IDs, but increased with increasing iron load above moderate according to the General Visual Rating Scale (Table 4, Figure 4). For 10 or less IDs, the mean inter-observer difference was +/- 1.5 IDs, and for more than 30 IDs counted by both analysts, it was +/- 6 IDs.

Volume of IDs

The volumes of IDs measured by the T2*W thresholding method were smaller than the volumes obtained by MCMxxxVI (Figure 5) for scans with moderate to high iron loads. For none or mild iron loads, rated according to the General Visual Rating Scale, the volumes provided by both methods were coincident. Both methods showed small systematic biases between observers, larger with MCMxxxVI (9 voxels) than with the T2*W thresholding method (2 voxels) (Figure

6). The 95% confidence interval of the measurements obtained by both methods varied in an interval of approximately 50% (T2*W thresholding) and 30% (MCMxxxVI) of the maximum volume of iron measured (Table 4).

For the intra-observer reliability tests done for MCMxxxVI, there was no difference between the first and second measurements (Figure 6) except for microbleeds in the brainstem and at the base of the cerebellum.

Visual Assessment

Microbleeds, using BOMBS

For microbleeds visually identified using BOMBS, the inter-observer difference in ID volume increased with increasing numbers of microbleeds, although the 95% confidence interval of the assessments only varied in an interval of ± 3 microbleeds and there was no bias, i.e. mean difference of -0.5 (Table 4). The variability was maximal in the areas of the deep structures, namely basal ganglia, internal/external capsules and thalamus, and minimal in lobar regions, namely subcortical white matter, cortex or grey/white matter junction (Figure 7).

IDs, using the Putaminal Iron Rating Scale, the General IDs Visual Rating Scale and Slice-per-Slice Visual Identification

Good results were obtained with MCMxxxVI without the visual assessment of the IDs on T2*W sequences. The average difference between the volume of IDs measured by MCMxxxVI with and without previous slice-per-slice visual identification was 38.74 voxels (Figure 8, Table 4) which is similar to the inter-observer difference.

The slices from each T2*W sequence where there was disagreement between observers in identifying IDs was counted. Table 5 shows the number of slices that differed between two

observers (inter-observer) and by the same observer (intra-observer). Differences of 6 or more slices in the inter-observer analysis occurred only 5 times amongst the 70 cases, representing only 7.1% of the total. In the intra-observer analysis, differences of 6 or more slices occurred only 3 times representing only 4.3% of the total. These small disagreements did not affect the visual rating according to the Putaminal ID Visual Rating Scale or to the General IDs Visual Rating Scale.

DISCUSSION

Both quantitative techniques for segmenting IDs in structural brain MRI have high reliability and repeatability. However, there are factors that limit their applicability to research and general clinical use. MCMxxxVI uses a clusterization method, MVQ, to segment the IDs. Taking advantage of the distributions of colors in an image, it produces segmentations of noticeably better quality than other clusterization techniques (21), but small details could still be lost. To achieve high accuracy during segmentation, the color of the IDs should produce good contrast with the surrounding voxels. In other words, there is the potential for very small and hypointense IDs to be missed by MCMxxxVI.

Our results suggest that the T2*W thresholding method could be used for counting microbleeds, but is not advisable for haemorrhages or other types of IDs; MCMxxxVI does not currently enable lesion counting. This is because large circumferences and sizes increase the error with the result that more manual post-processing editing is required to obtain optimum results. However, the thresholding method benefits from high sensitivity. Hence, accurate delineation of the hypointense areas in T2*W can be achieved, but, compared with MCMxxxVI, it also requires more manual editing and does not discern the areas where iron is associated with other minerals that also appear hypointense on T2*W, like calcium. Therefore parallel display of other scan

types is necessary to avoid misclassifying iron and calcium. Moreover, when segmenting IDs with unclear boundaries, the segmented region depends entirely on the threshold set by the observer. Here, setting of the threshold can be very variable and prone to bias depending on the observer's definition of the lesion, which in turn is likely to be affected by his/her neuroanatomical and neuroradiological knowledge and previous experience. Such subjective nature of thresholding and segmentation appears to have been a major contributory factor to the inter-observer variability seen when using the T2*W threshold-based method.

MCMxxxVI identifies areas where only iron is accumulated from those where other minerals are present by using combined information from two or more imaging modalities modulated in color space, in this case T2*W and FLAIR. It reduces the extent of subjective thresholding, but overestimates the boundaries of the IDs, mainly when the iron loads range from moderate to severe, making the results appear less favorable. Thus, although the results of our study show that both methods have high intra- and inter-observer reliability, more testing is required in patients with different pathologies at different ages, e.g. amyloid angiopathy, cavernous angioma and traumatic brain injury, using imaging data obtained from different scanners employing different scanning protocols prior to their using in the clinical setting.

In summary, both methods we tested for segmenting and measuring the volume of brain IDs are in agreement and have high reliability, but require a manual post-processing step to remove false positives that makes them observer dependant. Previous visual assessment of the IDs does not increase the reliability of these methods. The T2*W thresholding method is recommended for counting automatically the number of microbleeds, while MCMxxxVI is recommended for identifying putaminal IDs and distinguishing them from the areas where other minerals are present. More work is required to develop a reliable automatic method for use in

research and clinical practice that combines the advantages of both approaches and includes automated counting of IDs such as microbleeds.

References

1. Duckett S. Neuroradiology, In: Lea & Febiger editor. The Pathology of the Aging Human Nervous System.1 edition. Philadelphia: Lea & Febiger: 1991. 492 p.

2. Feldman HH, Maia LF, Mackenzie IRA, Forster BB, Martzke J, Woolfenden A. Superficial Siderosis: A Potential Diagnostic marker of Cerebral Amyloid Angiopathy in Alzheimer Disease. Stroke 2008;39:2894-2897.

3. Imaizumi T, Chiba M, Honma T, Niwa J. Detection of Hemosiderin Deposition by T2*-Weighted MRI After Subarachnoid Hemorrhage. Stroke 2003;34:1693-1698.

4. Slager UT,Wagner JA. The incidence, composition, and pathological significance of itracerebral vascular deposits in the basal ganglia. Neuropahology 1956;15:417-431.

5. Casanova MF,Araque JM. Mineralization of the basal ganglia: implications for neuropsychiatry, pathology and neuroimaging. Psychiatry Research 2003;121:59-87.

6. Haskins B,Leslie C. Basal Ganglia Mineralization in Psychiatry. Biological Psychiatry 1992;31:752-753.

7. van Harskamp NJ, Rudge P, Cipolotti L. Cognitive and social impairments in patients with superficial siderosis. Brain 2005;128:1082-1092.

8. Brass SD, Chen N, Mulkern RD, Bakshi R. Magnetic Resonance Imaging of iron Deposition in Neurological Disorders. Topics on Magnetic Resonance Imaging 2006;17:31-40.

9. Cordonnier C, Al-Shahi Salman R, Wardlaw J. Spontaneous brain microbleeds: systematic review, subgroup analyses and standards for study design and reporting. Brain 2007;1988-2003.

10. Forstl H, Krumm B, Eden S, Kohlmeyer K. What Is the Psychiatric Significance of Bilateral Basal Ganglia Mineralization. Biological Psychiatry 1991;29:827-833.

11. Szumowski J, Hayflick S, Gaarder K, Bas E, Schwarz E, Ergogmus D. Assessment of iron distribution in Hallevorden-Spatz Syndrome using phase imaging and relaxation rate measurements. In: Proceedings of the 17th International Society of Magnetic Resonance in Medicine, Honolulu, 2009. (abstract 3346).

12. Gootjes L, Teipel SJ, Zebuhr Y, Schwarz R, Leisinger G, Scheltens P, Moller HJ, and Hampel H. Regional Distribution of White Matter Hyperintensities in Vascular Dementia, Alzheimer's Disease and Healthy Aging. Dementia and Geriatric Cognitive Disorders 2004;18:180-188.

13. Prins ND, van Straaten ECW, van Dijk EJ et al. Measuring progression of cerebral white matter lesions on MRI: Visual rating and volumetrics. Neurology 2004;62:1533-1539.

14. Wen W, Sachdev P. The topography of white matter hyperintensities on brain MRI in healthy 60- to 64-year-old individuals. NeuroImage 2004;22:144-154.

15. Valdes-Hernandez MC, Ferguson KJ, Chapell F, Wardlaw JM. New multispectral data fusion technique in MRI for white matter lesion segmentation: method and comparison with thresholding in FLAIR images. European Radiology 2010;20:1684-1691.

16. Cordonnier C, Potter GM, Jackson CA et al. Improving inter-rater agreement about brain microbleeds: development of the Brain Observer MicroBleed Scale (BOMBS). Stroke 2009;40:94-99.

17. Jenkinson M, Bannister PR, Brady JM, Smith SM. Improved optimisation for the robust and accurate linear registration and motion correction of brain images. NeuroImage 2002;17:825-841.

18. Foley JD, van Dam A, Feiner SK, Hughes JF. Computer Graphics: Principles and Practice in C, 2nd. edition.Addison-Wesley Professional:1996. p. 1-1200.

19. Guillermaud R,Brady M. Estimating the bias field of MR images. IEEE Trans.on Medical Images 1997;16:238-251.

20. Bland JM,Altman DG. Statistical Methods for Assessing Agreement Between Two methods of Clinical Measurement. Lancet 1986;1:307-310.

21. Diaz Acosta, Beatriz. Experiments in Image Segmentation for Automatic US License Plate Recognition. MSc. Thesis. Virginia Polytechnic Institute and State University. Department of Computer Science. 18-6-2004, p. 104.

Tables

Table 1. Parameters for each structural sequence used in this study. The slice-gap was zero in all sequences.

Sequence	T1W	T2W FSE	T2W GRE (T2*W)	FLAIR FSE
TR/TE/TI (ms)	9.8/4/500	11320/104.9	940/15	9002/147.38/ 2200
Orientation	coronal	axial	axial	axial
Slice thickness (mm)	1.3	2	2	4
Bandwidth (KHz)	15.63	20.83	12.50	15.63
FOV (mm)	256 × 256	256 × 256	256 × 256	256 × 256
Measurement time	8min, 12s	3min, 35s	5min, 53s	5min, 25s
Number of slices	160	80	80	40
Matrix	192 × 192*	256 × 256	256 × 256	256 × 256

* Zero-filled to 256 × 256

Table 2. Statistical characteristics of the 70 subjects grouped according to the iron load rated by two visual scales. For both scales, the median value for the group was 1, the standard error for skewness was 0.287 and the standard error for kurtosis was 0.566.

Scale	Statistics of the sample distribution	Load	Number of subjects	Percent in the sample
Basal ganglia	70 subjects	0	14	20
(0 to 4)	Mean: 1.69	1	27	38.6
	Skewness: 0.522	2	8	11.4
	Kurtosis: -1.052	3	9	12.9
		4	12	17.1
Whole iron load	70 subjects	0	14	20
	Mean: 1.39	1	27	38.6
(0 to 3)	Skewness:0.235	2	17	24.3
	Kurtosis:-0.957	3	12	17.1

Table 3. Results of testing for any relationship between parenchymal hyper-intense lesions and CSF volume and iron load as rated by the two visual rating scales in 325 subjects using a Bonferroni corrected ANOVA test.

Independent variables in the regression tests	Visual Rating scale	*p*-value
Volume of the hyper-intense lesions	0 to 4 basal ganglia	0.475
	0 to 3 whole iron load	0.325
CSF volume	0 to 4 basal ganglia	0.212
	0 to 3 whole iron load	0.181

Table 4. Results of the inter- and intra-observer reliability tests, given in number of voxels per subject. The visual assessment was done on T2*W images.

Differences in:	Number of obser-vers	Method	Mean differ-ence (voxels)	95% Confidence Interval (± 2SD)	Min. (absolute value) (voxels)	Max. (absolute value) (voxels)
Number of IDs counted	2	Thresholding	3.898	± 14.781	0	23
Total volume of iron deposits	2	Thresholding	-2.277	± 57.649	0	138
	2	MCMxxxVI *	29.869	± 357.8325	0	636
	1	MCMxxxVI *	53.087	± 466.24	0	1536
	1	MCMxxxVI with and without prior visual assessment	38.746	± 391.526	0	933
Number of BMBs counted	2	Visual assessment (BOMBS)	-0.543	± 3.058	0	5
Number of slices with IDs	2	Visual assessment	2.143	± 4.902	0	12
	1	Visual assessment	1.143	± 2.285	0	9

* indicates that the method was applied after the images were visually assessed.

Table 5. Number of slices in which the visual assessment of IDs differed.

Slice difference	Inter-observer Frequency (number of slices)	Percent (out of 70)	Intra-observer Frequency (number of slices)	Percent (out of 70)
0	23	32.9	42	60
1	14	20	9	12.9
2	7	10	6	8.6
3	10	14.3	5	7.1
4	5	7.1	4	5.7
5	6	8.6	1	1.4
6	1	1.4	0	0
7	1	1.4	2	2.9
8	1	1.4	0	0
9	1	1.4	1	1.4
12	1	1.4	0	0
Total	70	100	70	100

Figure legends

Figure 1. Putaminal Iron Deposits Scale. T2*W axial slices that show, from left to right, a representative example of loads 1, 2, 3 and 4 respectively.

Figure 2. a) T2*W (left) and FLAIR (right) axial slices of a subject with a mineralised basal ganglia. b) Resulting fused image in the Red/Green color space.

Figure 3. CSF and WMH volumes per visually-rated iron load. These measurements were done in all 325 subjects. The iron load was rated according to two visual scales, one for putaminal iron deposits and the general rating scale for putaminal iron, microbleeds and old haemorrhages (Section 3).

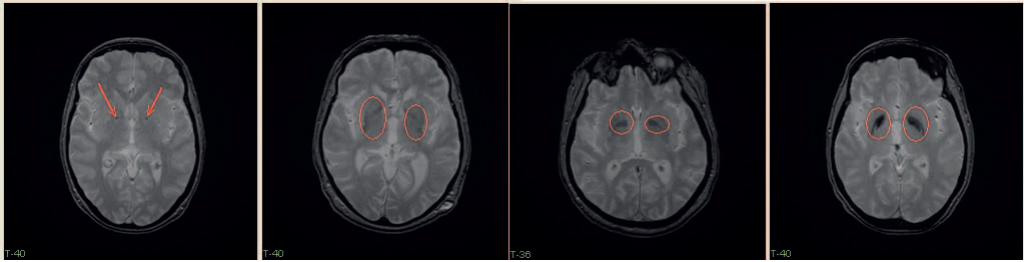
Figure 4. Bland-Altman plot of the results obtained by two analysts using the thresholding method to count the number of IDs in 70 subjects.

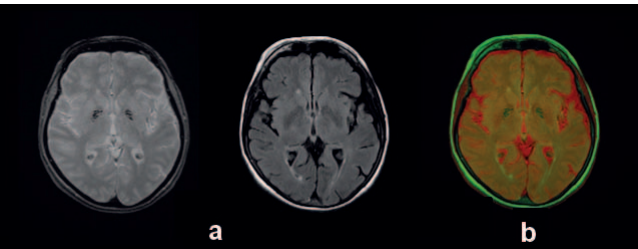
Figure 5. Bland-Altman plots. Inter-observer variability for IDs segmentation using MCMxxxVI and thresholding for T2*W sequences in 70 subjects, expressed in number of voxels per subject.

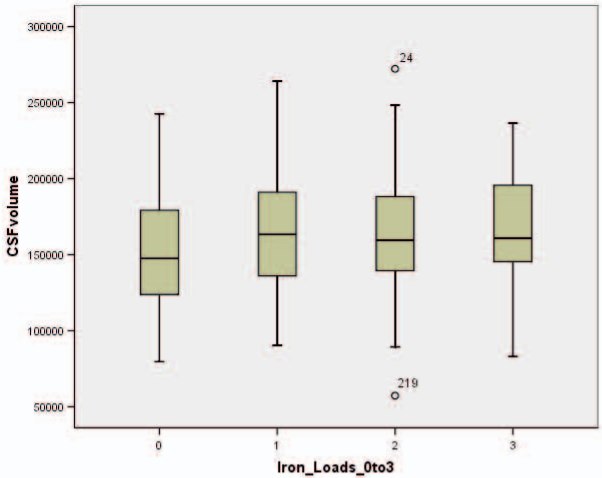
Figure 6. Bland-Altman plot. Intra- and inter-observer variability for IDs segmentation using MCMxxxVI in 70 subjects, expressed in number of voxels per subject.

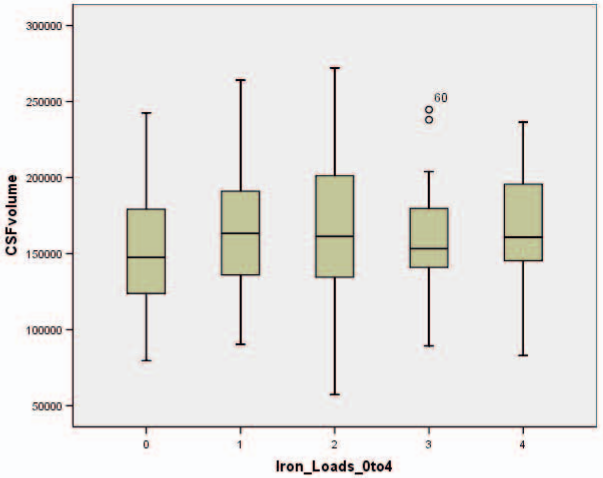
Figure 7. Bland-Altman plot of the inter-observer assessment of microbleeds using BOMBS in 70 subjects, expressed in number of brain microbleeds.

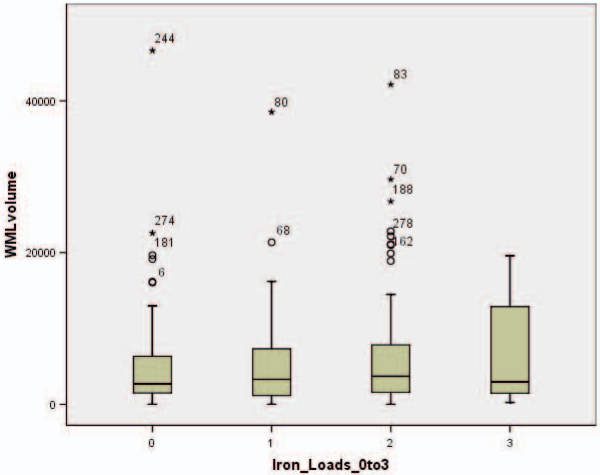
Figure 8. Bland-Altman plot. Reliability of MCMxxxVI with versus without concurrent visual assessment in the T2*W sequences in 70 subjects, expressed in number of voxels per subject).

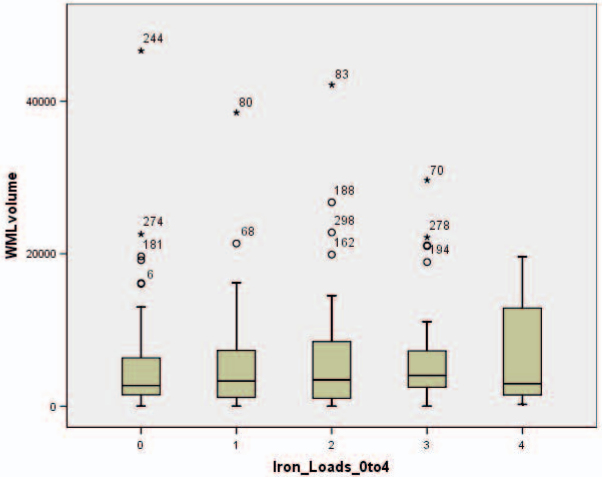

T-40
T-40
T-36
T-40

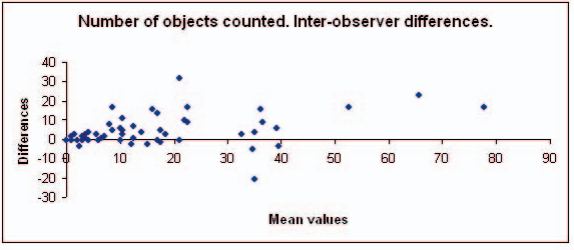

a
b

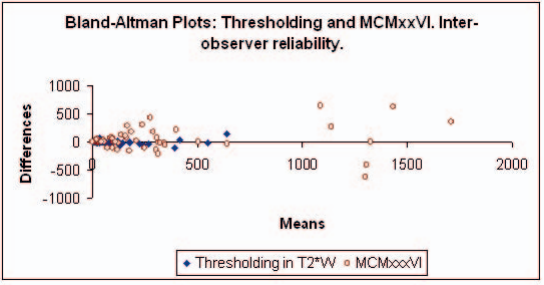

CSFvolume
300000
250000
200000
150000
100000
50000
24
219
0
1
2
3
Iron_Loads_0to3

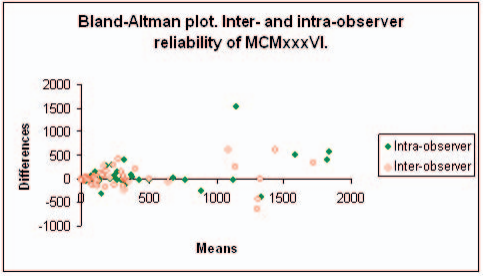

CSFvolume
300000
250000
200000
150000
100000
50000
60
0
1
2
3
4
Iron_Loads_0to4

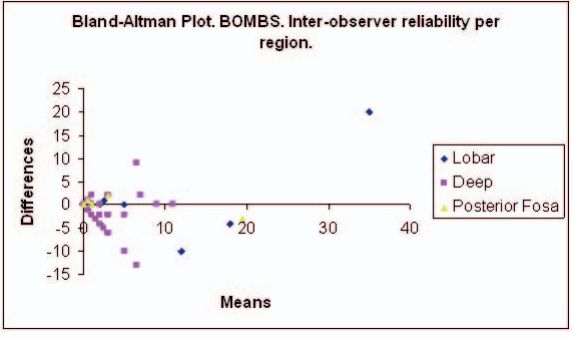

WMLvolume
40000
20000
0
244
83
80
70
188
274
278
68
181
162
6
0
1
2
3
Iron_Loads_0to3

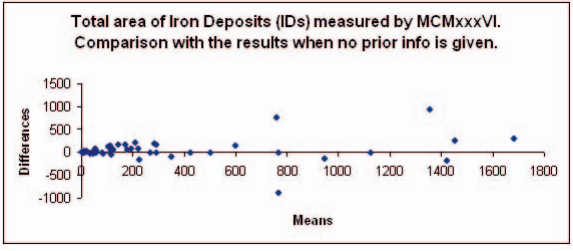

WMLvolume
40000
20000
0
244
83
80
70
188
274
298
278
68
181
162
194
6
0
1
2
3
4
Iron_Loads_0to4


Number of objects counted. Inter-observer differences.
Differences
40
30
20
10
0
-10
-20
-30
0
10
20
30
40
50
60
70
80
90
Mean values


Bland-Altman Plots: Thresholding and MCMxxVI. Inter-observer reliability.
Differences
1000
500
0
-500
-1000
0
500
1000
1500
2000
Means
Thresholding in T2*W
MCMxxxVI


Bland-Altman plot. Inter- and intra-observer reliability of MCMxxxVI.
Differences
2000
1500
1000
500
0
-500
-1000
0
500
1000
1500
2000
Means
Intra-observer
Inter-observer


Bland-Altman Plot. BOMBS. Inter-observer reliability per region.
Differences
Means
25
20
15
10
5
0
-5
-10
-15
0
10
20
30
40
Lobar
Deep
Posterior Fosa


Total area of Iron Deposits (IDs) measured by MCMxxxVI.
Comparison with the results when no prior info is given.
Differences
1500
1000
500
0
-500
-1000
0
200
400
600
800
1000
1200
1400
1600
1800
Means